Benke, I. (2020): Towards Design Principles for Trustworthy Affective Chatbots for Virtual Teams. Proceedings of the Twenty-Eighth European Conference on Information Systems (ECIS2020). Marrakesh, Morocco, June 15-17.

Institute of Information Systems and Marketing (IISM)
Kaiserstraße 89-93
76133 Karlsruhe - Germany
http://iism.kit.edu

Karlsruhe Service Research Institute (KSRI)
Kaiserstraße 89
76133 Karlsruhe – Germany
http://ksri.kit.edu

TOWARDS DESIGN PRINCIPLES FOR TRUSTWORTHY AFFECTIVE CHATBOTS FOR VIRTUAL TEAMS

Research in Progress

Benke, Ivo, Institute of Information Systems and Marketing (IISM), Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany, ivo.benke@kit.edu

Abstract

Virtual team communication has gained immense importance in recent years due to new work evolution and innovative IT-based communication tools. However, virtual teams face emotional obstacles within team communication. Affective chatbots can sense and understand human affective signals and leverage them to support the virtual team by increasing its emotional intelligence through behavioral and persuasive cues. However, through their capabilities such systems may also cause harm to individuals through addiction and increased vulnerability. Simultaneously, they experience higher distrust and skepticism. Therefore, affective chatbots require careful, ethical reflection on when and how to apply them in order to retain trustworthiness. In this paper, we present preliminary results of an ongoing design science research project developing design principles for affective chatbots with a specific emphasis on transparency and human autonomy. With our work we contribute theoretically with prescriptive design knowledge for the class of trustworthy affective chatbots in the context of virtual team communication. We, thereby, provide avenues towards a nascent design theory for this class of systems. Practically, our work supports providers of innovative IT-based communication tools in leveraging this knowledge and designing affective chatbots to help virtual teams in order to communicate more successfully under consideration of ethical principles.

Keywords: Affective Chatbots, Trustworthiness, Emotional Intelligence, Virtual Team, Design Science Research

1 Introduction

Virtual team communication has gained immense importance in recent years due to new work evolution (Frank et al., 2019). Today, over 50% percent of the working population in the United States work from remote (Forbes.com, 2019). Simultaneously, innovative IT-based communication tools like Slack or Microsoft Teams empower virtual team communication (Finnegan, 2019; Stoeckli et al., 2019). However, during communication virtual teams increasingly encounter serious problems like conflicts, breakdowns, or groupthink which highly disrupt the flow of communication.

All of these issues can be rooted back to effective management of team emotions (Barsade, 2002; Pitts et al., 2012). Since emotional information is limited during virtual communication, the capability to manage and process this limited information is crucial. Addressing that emotional capabilities are highly deterministic for team communication (Bartsch and Hübner, 2005), emotional intelligence (EI) represents the ability to sense, understand and regulate own and others emotions (Mayer et al., 2008). It is an influential factor on communication in virtual teams (Pitts et al., 2012) and research shows that weak EI may lead to communicative breakdowns (Bjørn and Ngwenyama, 2009). Besides the core functionality, innovative IT-based communication tools allow for integration of embedded third-party applications to increase productivity. Specifically, they have opened the gate for introducing chatbot applications into virtual team communication (Lechler et al., 2019). Therefore, we raise the question of '*why not using innovative chatbot applications in order to improve the emotional status of the team and support its emotional capabilities?'* Emotion-aware chatbots can sense and understand human emotions enabled through artificial intelligence

(AI) (Mensio et al., 2018). Through these capabilities they appear anthropomorphic and may increase the desire to interact with (McDuff and Czerwinski, 2018). We argue for affective chatbots as a further evolution which extends and applies emotion-aware capabilities in order to support virtual teams by enabling EI through behavioral and persuasive cues (Fogg, 2003; Nass et al., 1996). Through this approach team communication may be facilitated and team effectiveness may be increased (Pitts et al., 2012).

However, human emotions are very intimate and sensitive (Brave and Nass, 2009). AI-enabled detection and possible disclosure of innermost emotions is associated with personal vulnerability (Derlega, 1987; Moon, 2000). This leads to strong skepticism and distrust against systems that are able to expose emotions. Trust, in turn, is an important driver of acceptance and use of information systems. Since the application of such affective chatbots is very promising (Peng et al., 2019) through their massive application at the workplace and in private places and their potential to facilitate interpersonal communication, it is important that emotion-exposing systems regain trustworthiness. Trustworthiness is a characteristic of the trustee, which is informed by a set values and previous behaviors (Ben-Ner and Halldorsson, 2010). Because trustworthiness is at risk, it requires careful reflection on when and how to apply such systems, and ethical considerations about responsible usage (Dignum, 2017; McDuff and Czerwinski, 2018). Based on this foundation, we follow the demand of researchers like André et al. (2019) and Dignum (2017) for design principles for trustworthiness of AI-enabled systems. Essential, minimal requirements for operationalizing trustworthy design are transparency and human autonomy. However, research is scarce on how to implement transparency and human autonomy in order to design affective chatbots to retain trustworthiness. This leads us to the following research question:

> *How to design affective chatbots for virtual teams under consideration of transparency and human autonomy in order to increase their trustworthiness?*

In order to answer this research question, this study follows the design science research (DSR) paradigm adapting the publication schema of Gregor and Hevner (2013). The DSR paradigm is useful to address a real-world challenge and particularly suited to address the research gap of lack of design knowledge for trustworthy affective chatbots. On the foundation of EI theory (Mayer and Salovey, 1997), computers as persuasive actors paradigm (Fogg, 2003), the theoretical foundation of explanations (Gregor and Benbasat, 1999) and human agency theory (Bandura, 1989) we outline in this research-in-progress paper the first three steps of a DSR cycle. In the first cycle, we have assessed the defining characteristics for affective chatbots in virtual teams. Based on this prior work, in the second cycle, we focus on the development of transparent and autonomous design principles and instantiate them through a preliminary prototype which increases trustworthiness. Through our work we contribute with avenues towards a nascent design theory of concrete prescriptive guidance for this class of artifacts (e.g. trustworthy affective chatbots) (Gregor, 2006; Gregor and Hevner, 2013).

2 Conceptual Foundations

2.1 Virtual Teams and Emotional Intelligence

Virtual teams are comprised of individuals who work interdependently using computer-mediated communication to accomplish a shared objective (Martins et al., 2004). In contrast to face-to-face teams, virtual teams face unique obstacles to establish effective communication with regards to the lack of verbal and non-verbal cues in all forms of virtual technology, and ensuing problems like difficulties in conflict management or groupthink (Pitts et al., 2012). However, communication in teams has major influence on team effectiveness (Mathieu et al., 2008). At the other hand, emotions have strong effects on the individual and the team (Kelly and Barsade, 2001). In team interaction, EI plays an important role in order to deal with the limited amount of information (Pitts et al., 2012). EI is composed out of four constructs: the human ability of sensing, facilitating, understanding, and managing emotions (Mayer et al., 2008). Research shows that EI improves team communication and supports quality of interpersonal interaction in face-to-face teams (Melita Prati et al., 2003). Finally, EI is a strong predictor for job performance where social interaction exists (Joseph and Newman, 2010). Albeit positive effects, the development of EI in virtual settings is difficult due to the unique obstacles virtual team members face. First attempts of supporting

human EI through agents have been applied (Ivanović et al., 2014), however, EI support in virtual teams through innovative technology, like AI-enabled chatbots, remains scarce.

2.2 Affective Chatbots

Affective chatbots are based on the paradigm of affective computing (Pamungkas, 2017) which describes the extraction of human emotions by computers through sensors, feature extraction and signal derivation (Picard, 1995; Poria et al., 2017). Through advances in emotion recognition in conversation (Poria et al., 2019), chatbots are becoming increasingly able to distinguish emotions from team communication and are able to sense, understand and interpret human emotions (McDuff and Czerwinski, 2018). Such systems that are equipped with these abilities of sensing affective signals along with contextual information have been perceived more satisfying and activating (Bickmore and Cassell, 2001). With the creation of more natural and social interactions through emotional awareness together with anthropomorphic design components (Araujo, 2018; Feine et al., 2019; Rietz et al., 2019), they may support human-decision making, well-being and leverage this information for improving team interaction (Beck and Libert, 2017; Fogg, 2003; Reeves, 2000). Beyond the traditional application of chatbots as individual assistants, chatbots can be applied to multiparty interaction becoming a valid team member (Benke, 2019; Seeber et al., 2019). Therefore, we extend the ability of emotion-aware chatbots into affective chatbots which leverage emotional information for improving the EI of virtual teams.

2.3 Ethical Considerations and Trustworthiness of AI-Enabled Systems

Ethical consideration have been raised about AI-enabled systems and affective agents (EU, 2019; McDuff and Czerwinski, 2018). Several initiatives shed light on the threats and risks of such systems like the ethical guidelines for trustworthy AI (EU, 2019). They focused on the establishment of trustworthiness in order to allow for ethical conform application and usage of AI-enabled systems. Trustworthiness as system characteristic describes the trusting beliefs about system's competence, benevolence and integrity (McKnight et al., 2002) and influences trust into a system (McKnight et al., 2017). Trust is an important factor for a system's acceptance and usage for information systems (Lee and Choi, 2017) and has been proven to hold in the context of AI-enabled, intelligent agents as well (c.f. Wang and Benbasat (2005), Banks (2019)). To achieve trustworthiness two main aspects are considered in literature, technical robustness and ethical design (Mittelstadt, 2019). With regards to the operationalization, different suggestions have been made for trustworthy design of AI-enabled system. For example, in the case of anthropomorphic, intelligent agents, André et al. (2019) argued for a humane design, and Dignum (2017) identifies the need for responsible AI-enabled systems. Such endeavors in literature reveal the necessity of value-driven and trustworthy design for the case of affective chatbots. All of them pose the minimal requirement of transparency and human autonomy in order to fulfill ethical standards within the design of trustworthy AI-enabled systems. Transparency assures required understanding of the system's actions (Cramer et al., 2008). Autonomy is considered as self-determination of individuals which construct own goals and values, and are able to decide and act in their manner (Friedman and Nissenbaum, 1997). Following this approach, we derive the necessity to instantiate trustworthiness through the operationalization of these two constructs guaranteeing trustworthy design of affective chatbots.

3 Research Method

We conduct a DSR project following the DSR framework by Kuechler and Vaishnavi (2008) presented in Figure 1. The DSR paradigm seeks to design, build, and evaluate socio-technical artifacts that extend boundaries of descriptive knowledge in order to address unsolved problems in an innovative way or to solve known problems more effectively (Gregor and Hevner, 2013; Hevner et al., 2019, 2004). DSR studies, in general, follow process models, consisting of different phases like problem phase, suggestion phase, artifact development and evaluation as seen in Figure 1 (Kuechler and Vaishnavi, 2008). Following a DSR paradigm is a promising approach for our research endeavor since DSR focuses in particular on the development of useful artifacts (Baskerville et al., 2018; Hevner et al., 2004). In this paper, we only summarize the key findings from cycle 1 and put an emphasize on cycle 2.

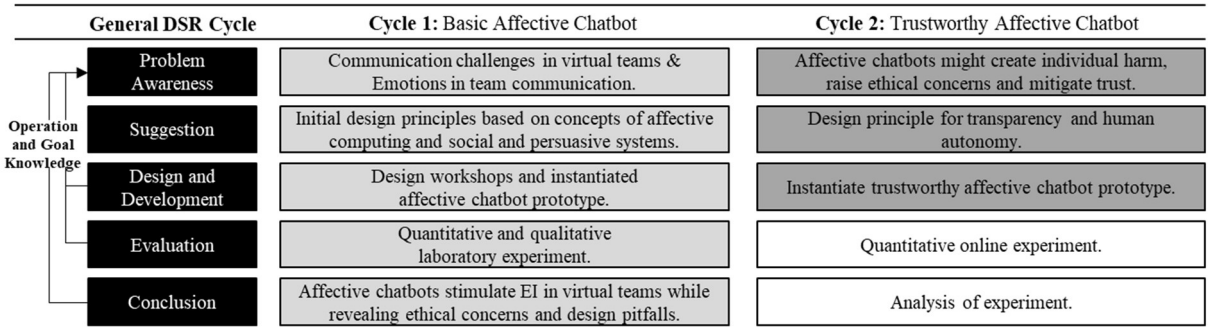

Figure 1. Overall DSR project.

Our work focuses on affective chatbots supporting EI in virtual teams by leveraging the paradigm of chatbots as social and persuasive actors. In cycle 1, we conducted initial design workshops with professionals and novices since there exists no foundational design knowledge for chatbots leveraging affective information in virtual teams beyond dyadic settings with focus on providing emotional support. Based on team models (Gilson et al., 2015) and characteristics of conversational systems (Feine et al., 2019; Fogg, 2003) 153 design sketches and subsequently three prototypes were elaborated. The evaluation revealed increased self-awareness and emotional perception, and improved consensus-seeking behavior and communication efficiency. The findings allowed us to formalize three design principles. However, we also documented pitfalls of the design such as a perceived lack of control (Wünderlich and Paluch, 2017), surveillance and indisposition (see also McDuff and Czerwinski, 2018; Mensio et al., 2018) which decreased trust. These drawbacks raised the need for a stronger trustworthy design (Dignum, 2017) in order to achieve the positive effects on EI while retaining trustworthiness.

The results of cycle 1 revealed an important need for trustworthy technology as future guidance for designing of affective chatbots. In cycle 2, we addressed this. Based on reviews of ethical guidelines for AI and the results of cycle 1, we draw the conclusion that the execution of human autonomy and transparency is necessary for trustworthy affective chatbots. To address these requirements, we propose one additional design principle based on the theoretical foundation of knowledge-based system explanations (Gregor and Benbasat, 1999) and human agency and control theory (Bandura, 2006; Frazier et al., 2011). Finally, we instantiated the design principle and developed a trustworthy affective chatbot prototype. In an online experiment the prototype will be evaluated in the future.

4 Conceptualization

4.1 Problem Awareness & Meta-Requirements

The first meta-requirement (MR1) refers to the system's ability to extract individual human emotional signals within a team during appearance without intercepting the communication flow. Emotions are a key influencing factor for team communication (Kramer, 1999; Ocker and Webb, 2009). Pitts et al. (2012) showed how they are impacting the communication quality and overall team effectiveness. At the same time, for most circumstances in virtual teams emotions cannot be transferred as they are in face-to-face teams through prevailing verbal and non-verbal cues (Martins et al., 2004). Summarized, virtual teams are faced with unique obstacles towards effective communication **(Issue 1)**. A system that aims to leverage emotional states in order to help the team, needs to be able to extract signals which might provide affective information of the users **(MR1.1)**. Team communication is characterized by quick succession of member contributions, a continuous flow of multiple changes of members' emotional states, and shifting of the team's attitude (Hepach et al., 2011). Therefore, a system which is adapting to human behavior during conversations, requires the ability to adopt new information fast and during appearance **(MR1.2)** (Lux et al., 2018). At the same time, team conversations are fragile (Bjørn and Ngwenyama, 2009). External factors like disturbances or distractions negatively influence the communication path, the way team members behave, and the communication outcome (Bartelt and Dennis, 2014). Consequently, a system which aims for improving virtual team communication needs to avoid them **(MR1.3)**.

MR1: *The system shall be able to extract team members' emotional signals during appearance without intercepting the flow of communication.*

The second meta-requirement (MR2) describes the ability to process, analyze and interpret the extracted emotional signals over the course of conversation. A key part of human-like AI is the understanding of human emotions (Poria et al., 2019). In order to stimulate the understanding and regulation of emotions, computers must understand human emotions first (Pentland, 2005). Therefore, the extracted emotional signals need to be analyzed and processed in order to allow a valid interpretation (**MR2.1**) (McDuff and Czerwinski, 2018). Two aspects are of higher importance. First, the information needs to be analyzed longitudinal over time due to conversational turns in team communication, which represent single units for emotional extraction, do not stand by their own and are not context-free (**MR2.2**) (Poria et al., 2019). Furthermore, achieving emotion understanding beyond individuals implies the understanding of multiple team members' emotions and therefore the combination of information of multiple sources (**MR2.3**).

MR2: *The system shall be able to analyze team members emotional signals over time.*

The third meta-requirement (MR3) refers to supporting the emotional management of virtual teams based on the extracted and analyzed emotional signals. Together with challenges like different cultural origin and characters, or unfamiliarity between team members, the lack of transfer capabilities of affective signals increases the complexity of emotion understanding. This aggravates the adequate reaction through emotion regulation which by itself is a complex process **(Issue 2)** (Adrianson, 2001). Emotions have impact on different outcomes of interaction, e.g. limited emotional understanding can lead to suboptimal decisions (Barsade, 2002) and lack of consensus creates instability (Barlow and Dennis, 2016), which might lead to communicational breakdowns (**Issue 3)** (Bjørn and Ngwenyama, 2009). Addressing those issues, a system should use the retrieved emotional information from the users in order to support the virtual team communication (**MR3.1**). Emotional breakdowns might originate in individual or team emotional conflicts. Therefore, a system needs to differentiate between the level of support (individual or team level) (**MR3.2**). Additionally, through inconsiderate disclosure of emotional information to the team in the wrong situation, social pressure to individuals may be created which a system should avoid (**MR3.3**).

MR3: *When the virtual team experiences lack of emotional capabilities, the system shall help the virtual team based on the collected emotional information either on the individual or team level.*

The fourth meta-requirement (MR4) targets the general design of form and function the system shows when interacting with team members. The ability of being emotion-aware allows for creating well-being, interacting in a more natural way, and providing more trustworthiness. A system supporting virtual teams with managing their emotions requires a specific setting and specific abilities (McDuff and Czerwinski, 2018). Since a team maintains specific social dynamics, an interacting system needs to follow clear rules to align with such dynamics (**MR4.1**) in order to become a social actor within the team (Nass and Moon, 2000). A machine interacting with humans is stronger accepted if it shows anthropomorphic appearance (**MR4.2**). Through becoming a social actor with anthropomorphic appearance, social relationships will be created. Social and emotional relationships require the system to adapt several factors in order to support the team in the best possible way like its social cues, its content or its role (**MR4.3**) (Fogg, 2003; Nass et al., 1996). The combination of those aspects forms a social entity which can seamlessly be integrated into social interaction of virtual teams.

MR4: *The system should integrate into the virtual team in a seamless and social way.*

The fifth meta-requirements (MR5) refers to the harm and ethical concerns that come along with emotion-aware, AI-enabled systems. Emotions lie at the core of human nature (Brave and Nass, 2009). Since they are very intimate and sensitive, humans are highly cautious on how to express real emotions (Hancock et al., 2008). Through their capabilities of interpreting and leveraging human emotions emotion-aware systems may cause severe harm to human psyche. Through knowledge on the current feeling systems can create addiction through empathetic behavior. This may even result in changes in behavior and personality. The system's knowledge may expose vulnerabilities of the human and can use it to manipulate and threaten the individual **(Issue 4)** (Mensio et al., 2018). Therefore, an emotion-aware system needs to carefully pay attention to these threats on human intimacy and vulnerability **(MR5.1)**. One of the main ethical problems is the creation of individual harm through AI-enabled systems **(Issue 5)** (Bostrom and Yudkowsky, 2011).

Due to obvious social threats, ethical considerations are necessary how to design and apply emotion-aware systems (**MR5.2**) (McDuff and Czerwinski, 2018). Furthermore, people tend to show resentment against new technologies as for example AI (EU, 2019). In the case of social expectations against a system are not matched these feelings are enforced. In consequence, considerations lead to stronger distrust **(Issue 6)**. However, trust is crucial for establishing a working relationship between the users and the system, and for letting users accept emotion-aware systems (**MR5.3**).

MR5: *The system should assure transparency and human autonomy during virtual team interaction.*

4.2 Design Principles

Based on the identified meta-requirements we derive four design principles (DP) for affective chatbots in virtual teams. Figure 2 depicts the mapping from issues, to meta-requirements, to design principles.

Following the paradigm of affective computing the system needs to be able to sense individual, affective verbal and nonverbal signals as well as contextual information (**MR1.1**) (Pentland, 2005; Picard, 1995). Since artificial, non-native interventions disturb the flow of team communication (**MR1.3**), a system which aims for pursing this objective and avoiding interceptions, should be as least immersive as possible. Simultaneously, the extraction of affective signals needs to happen during their appearance (**MR1.2**) which requires the system to process the information in real-time. Following MR2, the extracted signals should be analysed and aggregated to the team level to allow for team emotion interpretation (**MR2.1**). This process of analysis is conducted through fusion models which contain feature extraction, modelling of feature analysis structure and fusion of processed information (Poria et al., 2017). Emotions in conversations of virtual teams are dependent of precedent utterances and the context which requires systems to continuously extract and analyze emotional information (**MR2.2**). A fusion model, therefore, implements different utterances, and analyses the individual and the team level (**MR2.3**). Thus, we propose:

DP1: *Provide the affective chatbot with the ability of extracting and analyzing emotional signals from virtual team members using real-time behavioral data in a non-immersive way.*

A system should leverage its capabilities of emotion-awareness when teams require it in case of emotional communicative breakdowns (**MR3.1**). Increasing emotional understanding and supporting emotion management of one's own or others as main components of EI (Mayer and Salovey, 1997) may avoid or at least mitigate such processes or situations (Pondy, 1992; Xolocotzin Eligio et al., 2012). Chatbots are communicating via natural language which is more interactive and effective while being natural as well (Maes, 1994). Applications like Slack allow both for communicating in group channels as well as directly addressing of individuals which enables a multitude of affordances (Stoeckli et al. 2019) (**MR3.2**). When addressing multiple individuals within a team this can result quickly into a delicate situation which creates unpleasant and harmful situations through negative social dynamics (**MR3.3)** (Grudin, 1994) like blaming of individual team members (Behfar et al., 2008; Lowry et al., 2016). These dynamics create social pressure from the team to individual members (Pentland, 2005) which may lead to negative consequences like psychological harm. Aspects to prevent negative social pressure include education, role models but foremost a robust system design (Lowry et al., 2016) which provides a clear structure how to interact appropriately with stakeholders. Thus, we propose:

DP2: *Provide the affective chatbot with the ability to support emotional intelligence within the virtual team on the individual and team level based on the analyzed emotional information while avoiding harm to the individual.*

In order to support the team in the best possible manner, a system should integrate into the virtual team into seamless and social way. Humans tend to perceive machines as social actors (Nass et al., 1994). The human appearance by a chatbot may be achieved by social cues (Feine et al., 2019) like anthropomorphic attributes or behavior (Meza-de-Luna et al., 2019) (**MR4.2**). Such anthropomorphic design features may help to increase acceptance and the effect on EI support (McDuff and Czerwinski, 2018; Mou and Xu, 2017). This increases the natural interaction and well-being (Reeves, 2000). The emotion-awareness expands the abilities of a chatbot since it is able to adapt its design and social cues to the participants (Bian et al., 2016). A team conversation requires characteristics beyond traditional social cues towards more social interaction with conversational turns and states which allow for social behavior by the system

(**MR4.1 & 4.3**). Based on the paradigm of computers are persuasive actors (Fogg, 2003), chatbots can apply persuasive design features in order to enhance the positive effect on EI (Oinas-Kukkonen and Harjumaa, 2009). These include physical (e.g. facial expression) and psychological (e.g. humor) cues, social roles, the social dynamics and language style (Fogg, 2003). Thus, we propose:

DP3: *Provide the affective chatbot with anthropomorphic and persuasive design features.*

Emotion-aware system need to ensure ethical conform and trustworthy design. This requirement represents the core of this study and the focal design cycle. To make a step beyond purely functional affective chatbots it is desirable to achieve a trustworthy design through implementing transparency and human autonomy as minimal requirement. This avoids harm to stakeholders while possibly increasing trust and the intended effect on EI of the team members. Transparency is an integral aspect in order to assure required understanding of the system's actions (**MR5.2**) and to become trustworthy (**MR5.3**) (Cramer et al., 2008). It can be provided through system explanations which has been proven to increase trust as well (Gregor and Benbasat, 1999; Rader et al., 2018; Wang and Benbasat, 2005). Explanations vary in dimensions of content (reasoning, support, strategic, terminological), presentation format (automatically, user-invoked, or intelligent), and provision mechanism (text-based) (Gregor and Benbasat, 1999), which may be applied specifically in the context of affective chatbots in virtual teams. On the other hand, affective chatbots need to act on behalf of their human users (**MR5.1**). If chatbots do not act according to human motivations, human agency is at risk (Maedche et al., 2019). This autonomy of team members may be provided through human agency and control (Bandura, 1989; Frazier et al., 2011). To establish human agency, control mechanisms over a system can be established. Control mechanisms are categorized as behavioral and outcome mechanisms. They are provided through filter technologies to assess the nature of performing interventions by a system (Dabbish and Kraut, 2008). An operationalization might be the adaptation of timing, change of content or additional status information about parties (Dabbish and Kraut, 2004; McFarlane, 2002). Thus, we propose:

DP4: *Provide the affective chatbot with features ensuring transparency and autonomy through explanations and human agency and control mechanisms for virtual team members.*

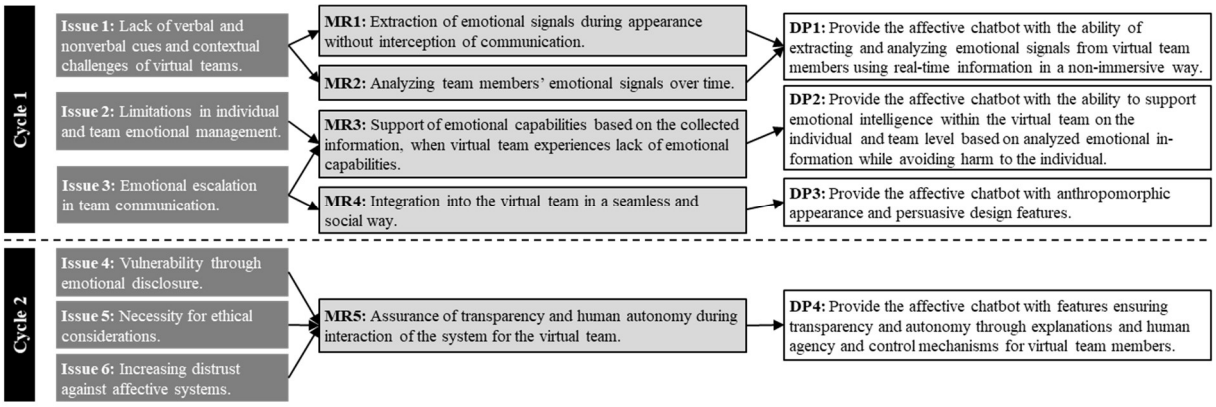

Figure 2. Issues, MRs and DPs for trustworthy affective chatbots in virtual teams.

4.3 Prototype Instantiation

DPs were instantiated into a DSR artifact following Kuechler and Vaishnavi (2008). Figure 3 presents a prototype of an affective chatbot in a team chat built after the example of Slack. The different DPs are translated into design features for both cycles. For the first design cycle, the artifact can extract information through advanced affective capabilities from text. Based on this information EI support actions are selected, and executed through design cues (see cycle 1 on the left). In cycle 2 we expand these DPs through design features of explanations, instantiated through an explanatory button and conversational explanations with the chatbot. Design features of autonomy are instantiated through control mechanisms like an on/off-switch (see cycle 2 on the right). After the instantiation, we are executing pilot-explorations with focus groups. Based on the initial results, we will conduct a large-scale online experiment to evaluate the effects of the DPs on transparency and autonomy in order to increase trustworthiness.

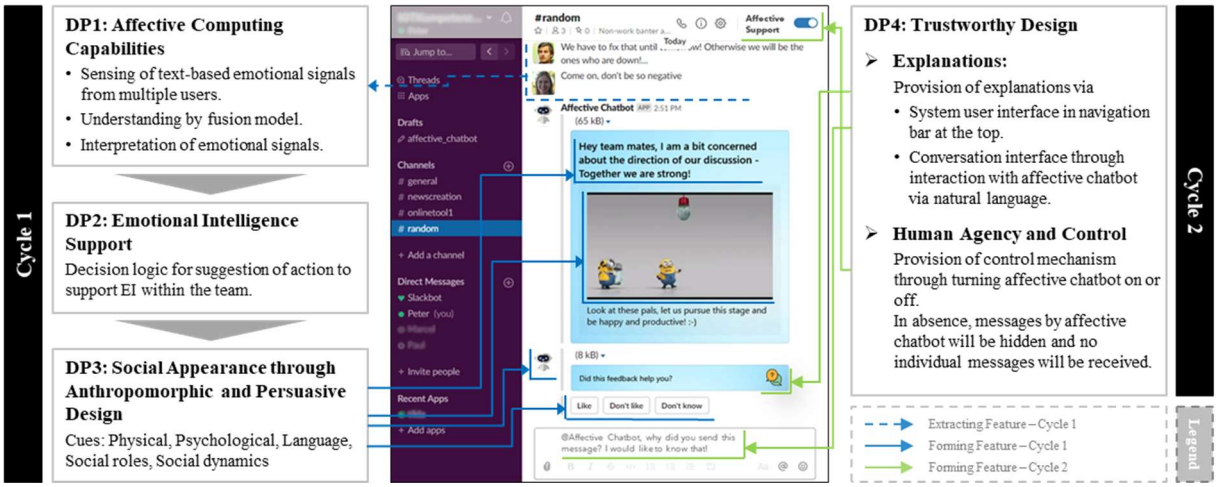

Figure 3. Prototype of affective chatbot with DPs 1-4 (recreated after Slack messenger example).

5 Conclusion and Expected Contribution

In this paper we present our ongoing DSR project on the design of affective chatbots for virtual teams with the focus on the introduction of trustworthiness through transparency and human autonomy. We discuss emotional obstacles of virtual teams using innovative communication technologies and the ethical concerns that arise with harmful application of affective chatbots. Building upon these issues, we propose MRs and DPs, and present a first prototype implementing the DPs. A technical risk and effectiveness evaluation strategy is planned according to Venable et al., (2016) as logical next step of our research. Nevertheless, to this paper several limitations apply. Due to its early phase, this project describes only preliminary MRs and DPs. These need to be refined throughout future research. Further, we focused on transparency and autonomy to achieve trustworthiness in affective chatbots. We are aware that these two construct are not exhaustive. However, we think that they are appropriate operationalizations since they represent core ethical principles (Jobin et al., 2019) while also being actionable in practice. Simultaneously, research is indicating their positive impact on trust which is highly important for acceptance of the system and the effect on EI.

In conclusion, this research is a step towards a nascent design theory (Gregor, 2006; Gregor and Hevner, 2013). We hope to provide valuable contribution to the body of prescriptive knowledge of affective chatbots for virtual teams, especially with the focus on trustworthy design (Dignum, 2017; EU, 2019). In practice, software providers of innovative IT-based communication tools can leverage this knowledge and design corresponding trustworthy affective chatbots to help virtual teams managing their emotions in order to communicate more successfully under consideration of ethical principles. Finally, through our DSR project, we aim to evolve the design of affective chatbots from simply successful and good into a humane and trustworthy user experience for the virtual team.

References

Adrianson, L., 2001. Gender and computer-mediated communication: Group processes in problem solving. Comput. Human Behav. 17, 71–94. https://doi.org/10.1016/S0747-5632(00)00033-9

André, E., Bayer, S., Benke, I., Benlian, A., ..., 2019. Humane Anthropomorphic Agents : The Quest for the Outcome Measure, in: Pre-ICIS Workshop on Values and Ethcis in AI. pp. 1–18.

Araujo, T., 2018. Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. Comput. Human Behav. 85, 183–189. https://doi.org/10.1016/j.chb.2018.03.051

Bandura, A., 1989. Human Agency in Social Cognitive Theory. Am. Psychol. 44, 1175–1184.

Banks, J., 2019. A perceived moral agency scale: Development and validation of a metric for humans

and social machines. Comput. Human Behav. 90, 363–371. https://doi.org/10.1016/j.chb.2018.08.028

Barlow, J.B., Dennis, A.R., 2016. Not As Smart As We Think: A Study of Collective Intelligence in Virtual Groups. J. Manag. Inf. Syst. 33, 684–712. https://doi.org/10.1080/07421222.2016.1243944

Barsade, S.G., 2002. The ripple effect: Emotional contagion and its influence on group behavior. Adm. Sci. Q. 47. https://doi.org/10.2307/3094912

Bartelt, V.L., Dennis, A.R., 2014. Nature and nurture: The impact of automaticity and the structureation of communication on virtual team behavior and performance. MIS Q. 38, 521–538. https://doi.org/10.25300/MISQ/2014/38.2.09

Bartsch, A., Hübner, S., 2005. Towards a Theory of Emotional Communication. CLCWeb Comp. Lit. Cult. 7.

Baskerville, R., Gregor, S., Baiyere, A., Hevner, A., Rossi, M., 2018. Design Science Research Contributions: Finding a Balance between Artifact and Theory. J. Assoc. Inf. Syst. 19, 358–376. https://doi.org/10.17705/1jais.00495

Beck, M., Libert, B., 2017. The Rise of AI Makes Emotional Intelligence More Important. Harv. Bus. Rev. 53, 1689–1699. https://doi.org/10.1017/CBO9781107415324.004

Behfar, K.J., Peterson, R.S., Mannix, E.A., Trochim, W.M.K., 2008. The Critical Role of Conflict Resolution in Teams: A Close Look at the Links Between Conflict Type, Conflict Management Strategies, and Team Outcomes. J. Appl. Psychol. 93, 170–188. https://doi.org/10.1037/0021-9010.93.1.170

Ben-Ner, A., Halldorsson, F., 2010. Trusting and trustworthiness: What are they, how to measure them, and what affects them. J. Econ. Psychol. 31, 64–79. https://doi.org/10.1016/j.joep.2009.10.001

Benke, I., 2019. Social Augmentation od Enterprise Communication Systems for Virtual Teams Using Chatbots, in: Proceedings of 17th European Conference on Computer-Supported Cooperative Work-Doctoral Colloquium. European Society for Socially Embedded Technologies (EUSSET).

Bian, Y., Yang, C., Guan, D., Xiao, S., Gao, F., Shen, C., Meng, X., 2016. Effects of pedagogical agent's personality and emotional feedback strategy on Chinese students' learning experiences and performance: A study based on virtual Tai Chi training studio. Conf. Hum. Factors Comput. Syst. - Proc. 433–444. https://doi.org/10.1145/2858036.2858351

Bickmore, T., Cassell, J., 2001. Relational agents: A model and implementation of building user trust. Conf. Hum. Factors Comput. Syst. - Proc. 396–403.

Bjørn, P., Ngwenyama, O., 2009. Virtual team collaboration: Building shared meaning, resolving breakdowns and creating translucence. Inf. Syst. J. 19, 227–253. https://doi.org/10.1111/j.1365-2575.2007.00281.x

Bostrom, N., Yudkowsky, E., 2011. The Ethics of Artificial Intelligence. Cambridge Handb. Artif. Intell. 1–20. https://doi.org/10.1017/CBO9781139046855.020

Brave, S., Nass, C., 2009. Emotion in Human–Computer Interaction 53–68. https://doi.org/10.1201/b10368-6

Cramer, H., Evers, V., Ramlal, S., Van Someren, M., Rutledge, L., Stash, N., Aroyo, L., Wielinga, B., 2008. The effects of transparency on trust in and acceptance of a content-based art recommender. User Model. User-Adapted Interact. 18, 455–496. https://doi.org/10.1007/s11257-008-9051-3

Dabbish, L., Kraut, R., 2008. Awareness displays and social motivation for coordinating communication. Inf. Syst. Res. 19, 221–238. https://doi.org/10.1287/isre.1080.0175

Dabbish, L., Kraut, R.E., 2004. Controlling Interruptions: Awareness Displays and Social Motivation for Coordination, in: Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW'04). pp. 182–191. https://doi.org/10.1287/isre.1080.0175

Derlega, V.J., 1987. Self-Disclosure: Inside or Outside the Mainstream of Social Psychological Research. J. Soc. Behav. Pers. 3, 27.

Dignum, V., 2017. Responsible Artificial Intelligence: Designing AI for Human Values. ICT Discov. 1–8.

EU, 2019. Ethics Guidelines for Trustworthy AI [WWW Document]. HLEG AI, Eur. Comm. URL https://ec.europa.eu/futurium/en/ai-alliance-consultation

Feine, J., Gnewuch, U., Morana, S., Maedche, A., 2019. A Taxonomy of Social Cues for Conversational

Agents. Int. J. Hum. Comput. Stud. 132, 138–161. https://doi.org/10.1016/j.ijhcs.2019.07.009
Finnegan, M., 2019. Collaboration 2019: Teams, Slack and what's coming [WWW Document]. computerworld.com. URL https://www.computerworld.com/article/3329540/collaboration-2019-teams-slack-and-whats-coming.html
Fogg, B., 2003. Computers as Persuasive Social Actors. Persuas. Technol. Using Comput. to Chang. What We Think Do 17, 269–274. https://doi.org/10.1016/S0749-3797(99)00093-8
Forbes.com, 2019. 10 Remote Work Trends That Will Dominate 2019.
Frank, M.R., Autor, D., Bessen, J.E., Brynjolfsson, E., Cebrian, M., Deming, D.J., Feldman, M., Groh, M., Lobo, J., Moro, E., Wang, D., Youn, H., Rahwan, I., 2019. Toward understanding the impact of artificial intelligence on labor. Proc. Natl. Acad. Sci. U. S. A. 116, 6531–6539. https://doi.org/10.1073/pnas.1900949116
Frazier, P., Keenan, N., Anders, S., Perera, S., Shallcross, S., Hintz, S., 2011. Perceived Past, Present, and Future Control and Adjustment to Stressful Life Events. J. Pers. Soc. Psychol. 100, 749–765. https://doi.org/10.1037/a0022405
Friedman, B., Nissenbaum, H., 1997. Software Agents and User Autonomy, in: Autonomous Agents.
Gilson, L.L., Maynard, M.T., Jones Young, N.C., Vartiainen, M., Hakonen, M., 2015. Virtual Teams Research: 10 Years, 10 Themes, and 10 Opportunities. J. Manage. 41, 1313–1337. https://doi.org/10.1177/0149206314559946
Gregor, S., 2006. The Nature of Theory in Information Systems. Manag. Inf. Syst. Q. 30, 611–642.
Gregor, S., Benbasat, I., 1999. Explanations from intelligent systems: Theoretical foundations and implications for practice. Manag. Inf. Syst. Q. 23, 497–530. https://doi.org/10.2307/249487
Gregor, S., Hevner, A.R., 2013. Positioning and Presenting Design Science for Maximum Impact. MIS Q. 37, 337–355. https://doi.org/10.2753/MIS0742-1222240302
Grudin, J., 1994. Groupware and Social Dynamics: Eight Challenges for Developers. Commun. ACM 37.
Hancock, J.T., Gee, K., Ciaccio, K., Lin, J.M., 2008. I'm sad you're sad: emotional contagion in CMC, in: Proceegs of CSCW '08: ACM Conference on Computer Supported Cooperative Work. pp. 295–298. https://doi.org/10.1145/1460563.1460611
Hepach, R., Kliemann, D., Grüneisen, S., Heekeren, H.R., Dziobek, I., 2011. Conceptualizing emotions along the dimensions of valence, arousal, and communicative frequency-mplications for social-cognitive tests and training tools. Front. Psychol. 2, 1–9. https://doi.org/10.3389/fpsyg.2011.00266
Hevner, A., vom Brocke, J., Maedche, A., 2019. Roles of Digital Innovation in Design Science Research. Bus. Inf. Syst. Eng. 61, 3–8. https://doi.org/10.1007/s12599-018-0571-z
Hevner, A.R., March, S.T., Park, J., Ram, S., 2004. Design Science in Information Systems Research. MIS Q. 28, 75–105. https://doi.org/10.2307/25148625
Ivanović, M., Radovanović, M., Budimac, Z., Mitrović, D., Kurbalija, V., Dai, W., Zhao, W., 2014. Emotional Intelligence and Agents, in: WIMS. pp. 1–7. https://doi.org/10.1145/2611040.2611100
Jobin, A., Ienca, M., Vayena, E., 2019. The global landscape of AI ethics guidelines. Nat. Mach. Intell. 1, 389–399. https://doi.org/10.1038/s42256-019-0088-2
Joseph, D.L., Newman, D.A., 2010. Emotional Intelligence: An Integrative Meta-Analysis and Cascading Model. J. Appl. Psychol. 95, 54–78. https://doi.org/10.1037/a0017286
Kelly, J.R., Barsade, S.G., 2001. Mood and emotions in small groups and work teams. Organ. Behav. Hum. Decis. Process. 86, 99–130. https://doi.org/10.1006/obhd.2001.2974
Kramer, R.M., 1999. TRUST AND DISTRUST IN ORGANIZATIONS: Emerging Perspectives, Enduring Questions. Annu. Rev. Psychol. 50, 569–598. https://doi.org/10.1146/annurev.psych.50.1.569
Kuechler, B., Vaishnavi, V., 2008. Theory Development in Design Science Research: Anatomy of a Research Project. Proc. Third Int. Conf. Des. Sci. Res. Inf. Syst. Technol. May 7-9, 1–15. https://doi.org/10.1057/ejis.2008.40
Lechler, R., Stoeckli, E., Rietsche, R., 2019. Looking Beneath the Tip of the Iceberg : the Two-Sided Nature of Chatbots and Their Roles for Digital Feedback Exchange. Proceeding ECIS 2019 1–17.
Lee, S.Y., Choi, J., 2017. Enhancing user experience with conversational agent for movie recommendation: Effects of self-disclosure and reciprocity. Int. J. Hum. Comput. Stud. 103, 95–

105. https://doi.org/10.1016/j.ijhcs.2017.02.005
Lowry, P.B., Zhang, J., Wang, C., Siponen, M., 2016. Why Do Adults Engage in Cyberbullying on Social Media? An Integration of Online Disinhibition and Deindividuation Effects with the Social Structure and Social Learning Model. Inf. Syst. Res. 27, 962–986. https://doi.org/10.1287/isre.2016.0671
Lux, E., Adam, M., Dorner, V., Helming, S., Knierim, M.T., Weinhardt, C., 2018. Live Biofeedback as a User Interface Design Element: A Review of the Literature. Commun. Assoc. Inf. Syst. https://doi.org/10.1016/j.ultsonch.2014.04.018
Maedche, A., Legner, C., Benlian, A., Berger, B., Gimpel, H., Hess, T., Hinz, O., Morana, S., Söllner, M., 2019. AI-Based Digital Assistants. Bus. Inf. Syst. Eng. 61, 535–544. https://doi.org/10.1007/s12599-019-00600-8
Maes, P., 1994. Agents that Reduce Work and Information Overload. Commun. ACM 37, 30–40. https://doi.org/10.1145/176789.176792
Martins, L.L., Gilson, L.L., Maynard, M.T., 2004. Virtual teams: What do we know and where do we go from here? J. Manage. 30, 805–835. https://doi.org/10.1016/j.jm.2004.05.002
Mathieu, J., Maynard, T.M., Rapp, T., Gilson, L., 2008. Team effectiveness 1997-2007: A review of recent advancements and a glimpse into the future. J. Manage. 34, 410–476. https://doi.org/10.1177/0149206308316061
Mayer, J.D., Roberts, R.D., Barsade, S.G., 2008. Human Abilities: Emotional Intelligence. Annu. Rev. Psychol. 59, 507–536. https://doi.org/10.1146/annurev.psych.59.103006.093646
Mayer, J.D., Salovey, P., 1997. What is Emotional Intelligence?, in: The Emotionally Intelligent Social Worker. pp. 10–23. https://doi.org/10.1007/978-0-230-36521-6_2
McDuff, D., Czerwinski, M., 2018. Designing emotionally sentient agents. Commun. ACM 61, 74–83. https://doi.org/10.1145/3186591
McFarlane, D.C., 2002. Comparison of four primary methods for coordinating the interruption of people in human-computer interaction. Human-Computer Interact. 17, 63–139. https://doi.org/10.1207/S15327051HCI1701_2
McKnight, D.H., Choudhury, V., Kacmar, C., 2002. Developing and validating trust measures for e-commerce: An integrative typology. Inf. Syst. Res. 13, 334–359. https://doi.org/10.1287/isre.13.3.334.81
McKnight, D.H., Lankton, N.K., Nicolaou, A., Price, J., 2017. Distinguishing the effects of B2B information quality, system quality, and service outcome quality on trust and distrust. J. Strateg. Inf. Syst. 26, 118–141. https://doi.org/10.1016/j.jsis.2017.01.001
Melita Prati, L., Douglas, C., Ferris, G.R., Ammeter, A.P., Buckley, M.R., 2003. Emotional Intelligence, Leadership Effectiveness and Team Outcomes. Int. J. Organ. Anal.
Mensio, M., Rizzo, G., Morisio, M., 2018. The Rise of Emotion-aware Conversational Agents, in: Companion Proceedings of the The Web Conference 2018. International World Wide Web Conferences Steering Committee. pp. 1541–1544. https://doi.org/10.1145/3184558.3191607
Meza-de-Luna, M.E., Terven, J.R., Raducanu, B., Salas, J., 2019. A Social-Aware Assistant to support individuals with visual impairments during social interaction: A systematic requirements analysis. Int. J. Hum. Comput. Stud. 122, 50–60. https://doi.org/10.1016/j.ijhcs.2018.08.007
Mittelstadt, B., 2019. Principles alone cannot guarantee ethical AI. Nat. Mach. Intell. https://doi.org/10.1038/s42256-019-0114-4
Moon, Y., 2000. Intimate Exchanges: Using Computers to Elicit Self-Disclosure From Consumers. J. Consum. Res. 26, 323–339. https://doi.org/10.1086/209566
Mou, Y., Xu, K., 2017. The media inequality: Comparing the initial human-human and human-AI social interactions. Comput. Human Behav. 72, 432–440. https://doi.org/10.1016/j.chb.2017.02.067
Nass, C., Fogg, B., Moon, Y., 1996. Can computers be teammates? Int . J . Hum. – Comput. Stud. 45, 669–678. https://doi.org/10.1006/ijhc.1996.0073
Nass, C., Moon, Y., 2000. Machines and Mindlessness: Social Responses to Computers. J. Soc. Issues 56, 81–103. https://doi.org/10.1111/0022-4537.00153
Nass, C., Steuer, J., Tauber, E.R., 1994. Computers are social actors, in: CHI 19. pp. 72–78. https://doi.org/10.1145/259963.260288

Ocker, R.J., Webb, H., 2009. Communication structures in partially distributed teams: The importance of inclusiveness, in: 15th Americas Conference on Information Systems 2009, AMCIS 2009. pp. 3231–3240.

Oinas-Kukkonen, H., Harjumaa, M., 2009. Persuasive Systems Design: Key Issues, Process Model, and System Features. Commun. ACM 24, 485–500.

Pamungkas, E.W., 2017. Emotionally-Aware Chatbots: A Survey, in: Proceedings of ACM Conference (Conference'17).

Peng, Z., Kim, T., Ma, X., 2019. GremoBot: Exploring emotion regulation in group chat, in: Proceedings of the ACM Conference on Computer Supported Cooperative Work and Social Computing. pp. 335–340. https://doi.org/10.1145/3311957.3359472

Pentland, A., 2005. Socially aware computation and communication. Proc. Seventh Int. Conf. Multimodal Interfaces, ICMI'05 199. https://doi.org/10.1145/1088463.1088466

Picard, R.W., 1995. Affective Computing. MIT Press 1–16. https://doi.org/10.1007/BF01238028

Pitts, V.E., Wright, N.A., Harkabus, L.C., 2012. Communication in Virtual Teams : The Role of Emotional Intelligence. J. Organ. Psychol. 28, 2046–2054. https://doi.org/10.1016/j.chb.2012.06.001

Pondy, L.R., 1992. Reflections on Organizational Conflict. J. Organ. Behav. 13, 257–261.

Poria, S., Cambria, E., Bajpai, R., Hussain, A., 2017. A review of affective computing: From unimodal analysis to multimodal fusion. Inf. Fusion 37, 98–125. https://doi.org/10.1016/j.inffus.2017.02.003

Poria, S., Majumder, N., Mihalcea, R., Hovy, E., 2019. Emotion Recognition in Conversation: Research Challenges, Datasets, and Recent Advances. IEEE Access 7, 100943–100953. https://doi.org/10.1109/access.2019.2929050

Rader, E., Cotter, K., Cho, J., 2018. Explanations as mechanisms for supporting algorithmic transparency. Conf. Hum. Factors Comput. Syst. - Proc. CHI 2018 1–13. https://doi.org/10.1145/3173574.3173677

Reeves, B., 2000. The Benefits of Interactive Online Characters. Cent. Study Lang. Information, … 1–11.

Rietz, T., Benke, I., Maedche, A., 2019. The Impact of Anthropomorphic and Functional Chatbot Design Features in Enterprise Collaboration Systems on User Acceptance, in: Proceedings of the 14th International Conference on Wirtschaftsinformatik.

Seeber, I., Bittner, E., Briggs, R.O., de Vreede, T., de Vreede, G.-J., Elkins, A., Maier, R., Merz, A., Oeste-Reiß, S., Randrup, N., Schwabe, G., Söllner, M., 2019. Machines as Teammates: A Research Agenda on AI in Team Collaboration. Inf. Manag. 103174. https://doi.org/10.1016/j.im.2019.103174

Stoeckli, E., Dremel, C., Uebernickel, F., Brenner, W., 2019. How affordances of chatbots cross the chasm between social and traditional enterprise systems, Electronic Markets. Electronic Markets. https://doi.org/10.1007/s12525-019-00359-6

Venable, J.R., Pries-Heje, J., Baskerville, R., 2016. FEDS: A Framework for Evaluation in Design Science Research. Eur. J. Inf. Syst. 25, 77–89. https://doi.org/10.1057/ejis.2014.36

Wang, W., Benbasat, I., 2005. Trust in and Adoption of Online Recommendation Agents. J. Assoc. Inf. Syst. 6, 72–101. https://doi.org/10.1016/j.jsis.2007.12.002

Wünderlich, N. V., Paluch, S., 2017. A Nice and Friendly Chat With a Bot: User Perceptions of AI-based Service Agents. Proc. Int. Conf. Inf. Syst. 1–11.

Xolocotzin Eligio, U., Ainsworth, S.E., Crook, C.K., 2012. Emotion understanding and performance during computer-supported collaboration. Comput. Human Behav. 28, 2046–2054. https://doi.org/10.1016/j.chb.2012.06.001